Postoperative neonatal mortality prediction using superlearning

Jennifer N. Cooper, PhD[1,2]; Peter C. Minneci, MD, MHSc[1,3,4]; Katherine J. Deans, MD, MHSc[1,3,4]

[1]Center for Surgical Outcomes Research and the Center for Innovation in Pediatric Practice, The Research Institute at Nationwide Children's Hospital, [2]Department of Pediatrics, The Ohio State University College of Medicine,[3]Department of Surgery, Nationwide Children's Hospital, [4]Department of Surgery, The Ohio State University College of Medicine

Introduction

- With the current availability of large electronic medical datasets, it is possible to develop robust clinical prediction models that enable clinicians to accurately estimate individual patients' risks of morbidity and mortality
- Traditionally, logistic regression models have been used to derive these predictions
- Logistic regression imposes stringent, parametric constraints on the relationship between predictors and the probability of an outcome
  - $\log(p/(1-p))=\beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_pX_p$
- Superlearning is an ensemble machine learning method for selecting via cross-validation the optimal algorithm among all weighted combinations of a set[1]
- Superlearning relies on an analyst-specified collection of algorithms and performance measure (e.g. squared difference between observed and predicted outcomes)
- The objective of this study was to develop and validate a clinical prediction model for 30-day postoperative mortality in neonates using superlearning

Methods

- Used 2012-14 National Surgical Quality Improvement Program-Pediatric data
- Patients treated in 2012-13 formed the development sample (N=6499, 3.6% mortality), and those treated in 2014 formed the validation sample (N=3552, 3.8% mortality)
- Used 211 preoperative predictors, 14 algorithms and 10-fold cross validation
  - 2 stepwise logistic regression models, 3 penalized logistic regression models, 2 generalized boosted regression models, 5 random forest models, and 2 classification tree models
- Repeated analysis after screening out predictors with p>0.20 in bivariate analysis
- Examined discrimination (AUROC) and calibration (calibration intercept and slope) of superlearner and all constituent algorithms in both the development and validation datasets

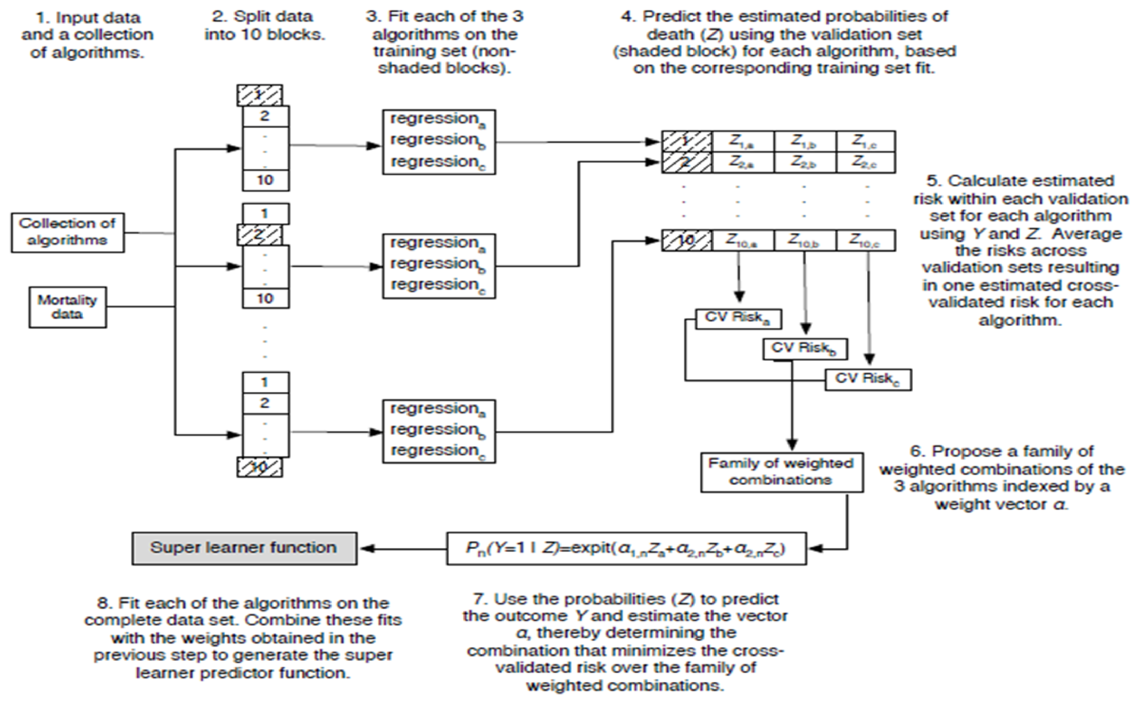

Figure 1. Super Learner Algorithm

From van der Laan & Rose, *Targeted Learning,* 2011

Table 1. Patient characteristics at the time of surgery

Characteristic	Patients treated in 2012-13		Patients treated in 2014	
	Survived 30 days (N=6267)	Died within 30 days (N=232)	Survived 30 days (N=3417)	Died within 30 days (N=135)
Age at surgery (days)	16 (3-53)	14 (7-28)	18 (3-61)	16 (6-32)
Gestational age at surgery (weeks)	39 (36-41)	33 (29-38)	39 (36-42)	34 (29-39)
Female	2405 (38.4)	103 (44.4)	1383 (40.5)	68 (50.4)
Race				
White	4041 (64.5)	131 (56.5)	2218 (64.9)	65 (48.1)
Black	968 (15.4)	53 (22.8)	589 (17.2)	36 (26.7)
Asian	136 (2.2)	5 (2.2)	88 (2.6)	3 (2.2)
Other/Unknown	1122 (17.9)	43 (18.5)	522 (15.3)	31 (23.0)
Weight at surgery (kg)	3.09 (2.43-3.74)	1.60 (1.06-2.92)	3.13 (2.48-3.82)	2.16 (1.09-3.32)
Ventilator dependent	1503 (24.0)	201 (86.6)	807 (23.6)	110 (81.5)
BPD or chronic lung disease	978 (15.6)	57 (24.6)	574 (16.8)	40 (29.6)
Oxygen support	1656 (26.4)	169 (72.8)	838 (24.5)	91 (67.4)
Structural pulmonary/airway abnormality	814 (13.0)	54 (23.3)	411 (12.0)	31 (23.0)
Esophageal, gastric, or intestinal disorder	3897 (62.2)	156 (67.2)	2169 (63.5)	88 (65.2)
Any cardiac risk factor	2339 (37.3)	123 (53.0)	1388 (40.6)	94 (69.6)
Structural CNS abnormality	1129 (18.0)	37 (15.9)	592 (17.3)	24 (17.8)
Open wound	732 (11.7)	18 (7.8)	332 (9.7)	11 (8.1)
Nutritional support	2691 (42.9)	160 (69.0)	1477 (43.2)	96 (71.1)
Hematologic disorder	784 (12.5)	78 (33.6)	481 (14.1)	59 (43.7)
Congenital malformation	2855 (45.6)	83 (35.8)	1804 (52.8)	60 (44.4)
Most common principal procedures				
Pyloromyotomy	687 (11.0)	0 (0)	410 (12.0)	0 (0)
Creation of VP shunt	362 (5.8)	3 (1.3)	204 (6.0)	5 (3.7)
Repair of large omphalocele or gastroschisis	329 (5.2)	11 (4.7)	179 (5.2)	6 (4.4)

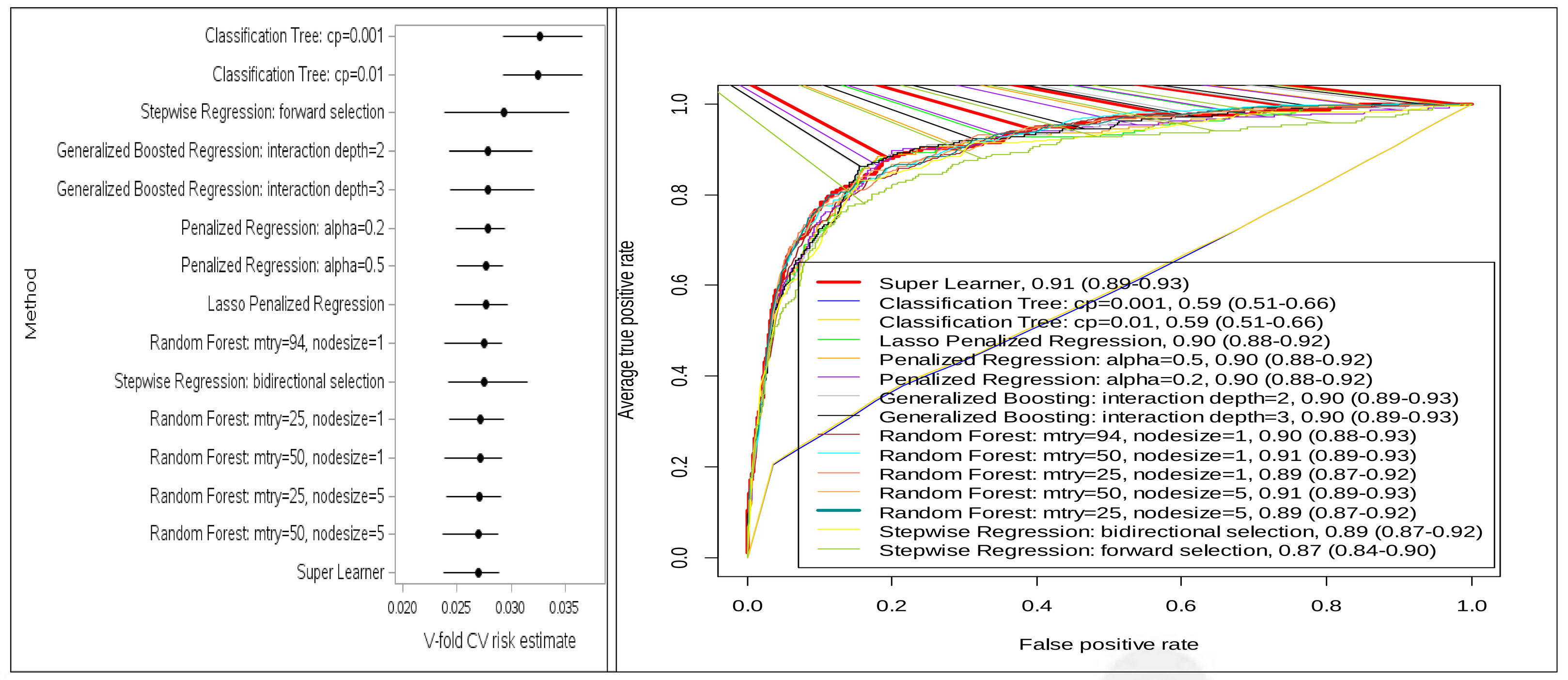

Figure 2. Cross-validated mean-squared error and AUROC of super learner and each candidate algorithm

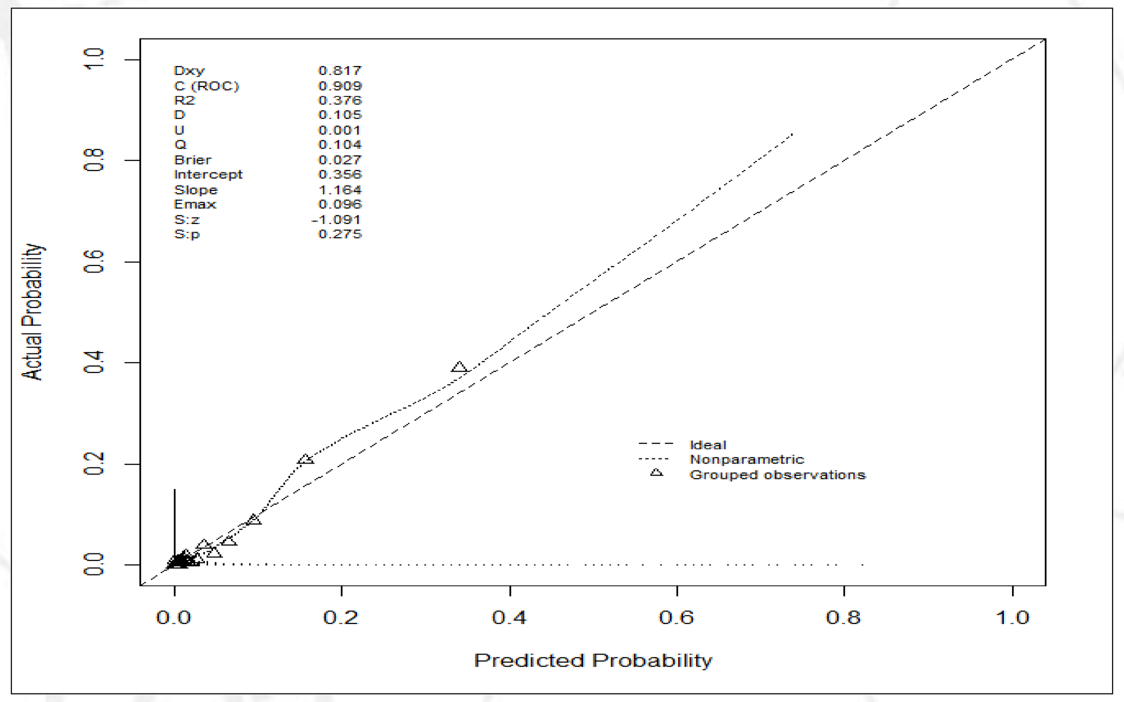

Figure 3. Calibration Plot

- The super learner showed good discrimination but poor calibration in the validation sample:
  - AUROC: 0.87 (95% CI 0.83-0.90)
  - Calibration: U statistic 0.064, p<0.001

Results

Table 2. Mortality predictions obtained from super learner for two different patient profiles

	Example low risk patient	Example high risk patient
Age at surgery (days)	28	37
Gestational age at surgery (weeks)	44	35
Gender	Male	Female
Race	White	White
Weight at surgery (kg)	3.97	1.15
Comorbidities	Esophageal, gastric, or intestinal disorder Congenital hypertrophic pyloric stenosis	Premature (25-26 completed weeks) Ventilator dependent BPD or chronic lung disease Supplemental oxygen support Esophageal, gastric, or intestinal disorder TPN or feeding tube Septic shock IV inotropic support Blood transfusion within 48h preop Obstructive apnea of newborn
Principal procedure	Pyloromyotomy	Exploratory laparotomy
Estimated mortality risk	0.0025	0.88

- Results were similar when superlearning was performed after variable screening:
  - AUROC: Development sample 0.91 (95% CI 0.89-0.93),
    Validation sample 0.87 (95% CI 0.83-0.91)
  - Calibration: Development sample U statistic -0.0004, p=0.09,
    Validation sample U statistic 0.068, p<0.001

Conclusions

- Superlearning provided improved or equivalent accuracy when compared to individual regression and machine learning algorithms for the prediction of neonatal surgical mortality but showed poor calibration in a validation sample
- Though many studies have reported algorithms for predicting neonatal mortality after particular types of surgical procedures, the vast majority have used main effects logistic regression modeling
- The Super Learner offers a flexible alternative to other non-parametric methods because it can include as many candidate algorithms as desired, and will perform at least as well as the best individual algorithm in its library
- Super learning should be considered for prediction in large datasets when complex mechanisms make parametric modeling assumptions unrealistic
- Although the Super Learner will perform no worse than the best constituent algorithm in a training dataset, there is no guarantee it will perform well in a validation dataset
  - Poor calibration in our validation data may have been due to 11 new hospitals joining the NSQIP-Pediatric program in 2014

References

1. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol* 2007;6:25.
2. Polley EC, van der Laan MJ. SuperLeaner: Super Learner Prediction. 2014; version 2.0-15: Available from: https://github.com/ecpolley/SuperLearner.

Acknowledgements

The American College of Surgeons National Surgical Quality Improvement Program and the hospitals participating in the ACS NSQIP Pediatric are the source of the data used herein; they have not verified and are not responsible for the statistical validity of the data analysis or the conclusions derived by the authors.